Network Working Group
Request for Comments #146
NIC 6742

P.M. Karp, MITRE
D.B. McKay, IBM
D.C. Wood, MITRE
12 May 1971

Categories: D.4, D.7
Obsoletes: none
Updates: none

Views on Issues Relevant to Data Sharing on Computer Networks

Introduction

The formation of a committee to address the problems of achieving data sharing on the ARPA Network, as suggested by Arie Shoshani (RFC #140) is desirable at this point of network development. We concur with Shoshani's ideas (presented in an introductory paper to the network data sharing meeting scheduled for Tuesday, May 18) and believe that purpose of the committee should be —

a) to classify the issues involved and to propose various approaches;

b) to integrate the hitherto independent network activities that address problems in the area of data sharing, and;

c) to set up and coordinate appropriate experiments to test the services developed and to evaluate alternative approaches.

This position paper is intended to augment Shoshani's as a basis for discussion at the data sharing meeting. No attempt is made to discuss specific means of implementation since many approaches to data handling problems are possible and have been proposed. Rather, our viewpoint on what the committee's role should be in giving some cohesion to various existing implementations is presented.

Our Views

One approach to achieving data sharing on the ARPA Network can be thought of as having three stages, which roughly correspond to the modes of use or operation. Within each stage are various levels of development required to get to the next stage. This development is not necessarily sequential. A description of the three stages follows.

Stage 1: Data handling services are provided at various Hosts. The user talks directly to the serving Host (via TELNET or by addressing a known socket) to explicitly access the service. This mode of operation corresponds to Bhushan's category of "direct" usage (RFC #114). The data services provided by the serving Host range from simple ones, such as White's file transfer system (RFC #122) to sophisticated systems such as the CCA's data machine (NIC 5791 and 6706).

Stage 2: The user has access to an intermediate process or data control facility* that routes his requests for a particular data service to the serving system. The user must explicitly identify the data services to be used. This mode of operation corresponds to Bhushan's category of "indirect" access. The data control facility provides the necessary control commands, data transformations, and accessing methods. A single request would include the use of several interacting services. For example, Heafner's Data Reconfiguration Service (RFC #138) could be used in conjunction with the use of CCA's data machine.

*The data control facility is not necessarily located at his local Host. Such a facility may exist on from one to all Host (i.e., ranging from centralized to completely distributed).

Stage 3: The user treats the network as a single resource and is unconcerned with the location of the services, data files, etc. All references are by name. In this mode of operation, the data control facility can function as a referral center for data service requests by using the most appropriate data service available and by automatically combining the use of several services that may be needed to satisfy a request. For example, data could be retrieved from several files, each managed by a different data management system. The data control facility must be cognizant of the location of data files, their structure, data management system capabilities, etc.

Some approaches to the design of the data control facility have been suggested by Shoshani, notably the integrated data management system (IDMS) and the unified data management system (UDMS). The notion of the network machine (RFC #51) is closest to the capabilities one would see in Stage 3.

Relevant Areas of Development

The data control facility can range anywhere from a simple interface to an intelligent front-end processor to a network-wide referral system. In any case, a common means is desirable for handling applications such as file transfer, on-line update and retrieval of data, information gathering and reporting, and program access to data. To attain this end, a few of the areas in which developments will be required include:

a) a data description language, permitting the user to define the physical structure of files, to define logical files, and to categorize data fields for name referencing. The language should be designed to facilitate the resolution of physical discrepancies in data and file structures. The user should be able to superimpose logical restructuring of data without any change in the physical structure.

b) a control or access language that can be mapped into various data management languages. Considered here is Shoshani's suggested two-level approach with perhaps a meta-language implementation to facilitate conversions among already existing languages.

c) methods for managing and merging distributed data, search mechanisms for file directories, error recovery techniques, etc.

Independent ARPA Network activities that in effect constitute Stage 1 have touched on these areas and should be incorporated into the overall data sharing scheme such that all of the isolated pieces are compatible. For example,

a) the data reconfiguration service (RFC #138) would be invoked by the data control facility whenever data transformations are required.

b) the file transfer protocol (RFC #114, #122) should be consistent with other data handling services.

c) CCA's data machine should be a subset or part of any data control facility. The network data language and set of data management services that they plan to implement can perhaps be adopted network-wide.

d) the network machine concept (RFC #51) for defining the program and data environments should be resurrected. The data control facility should be a subset of a network machine architecture.

Some other relevant topics include NIL (RFC #51), DEL (RFC 5), the notion of MYLOCAL n, YOUR LOCAL n, and STANDARD n (RFC #42), user level protocol objectives as described in RFC #76 and #91.

Experimentation and Testing

As data services are developed on the network, a coordinated effort is desirable

a) to exercise individual implementations to see if they work, both alone and in conjunction with other data services, and

b) to evaluate alternative approaches.

Some examples of experimentation to test data services follow:

1. File Transfer Protocol

   The file transfer protocol should be used to manipulate data files controlled by various systems.

2. Data Transfer to Data Computer

   The ability to transfer existing data bases and their structures onto the data computer should be demonstrated.

3. Data Restructuring

   The ability to define logical restructuring of data for users needs which would be accessible by name should be demonstrated. The original physical structure would be maintained.

4. Data Transformation

   The ability to access various data management systems on the network without the user being concerned with the data transformation involved should be demonstrated. Necessary calls to forms available on the Data Reconfiguration Service should be handled automatically and should be transparent to the user.

5. Data Consistency

   Problems of maintaining consistency when duplicate copies of a data file exist and updates to the file are made should be investigated. Automatic use of file transfer protocol and DRS to generate new duplicate copies should be included.

6. Data Privacy

   Access controls for privacy of data files in the network environment should be designed and evaluated. This includes controls on parts of distributed files.

Our recommendation is that the committee on data sharing be responsible for coordinating development in these areas, for attempting to maintain consistency among data services, and for testing services in a series of experiments as they are implemented.